



AI Feature Spec & PRD

✦ Draft with AI: a host gives the Studio a granth, and gets well-pitched, explained, checked questions to approve

Jain Sanyam Kamlesh, Product Manager & builder Status: **shipped to the live Studio as Beta, 14 Oct 2026**

Built with Claude (Anthropic)

1. Feature overview

Feature name: ✦ Draft with AI **Beta**, also called AI question drafts.

One line: the host uploads a source book (PDF), picks the pages and the mix of easy, medium and hard questions, and the AI drafts questions in that round's own format. Each draft carries its page number, a line from the book, an explanation, and the correct ways to type the answer. A second AI checks every answer blind, and the host approves each one before anything is added.

Where it appears: Studio → Content tab, beside + *New item* and *Add many at once*. This is step 3 of the flow in File 01 (section 3); see File 04, Figure 4.1.

Before

The host reads the granth, writes each question, invents three believable wrong options, and types it all in: **about an hour per show**. No explanations (0 of 15 in the 19 Sep game), no difficulty labels, and only one accepted spelling per typed answer. 53 of 145 "wrong" answers that night were actually right.

After

The host picks pages 41–50 and asks for 4 easy, 4 medium and 2 hard. **About a minute later** (65 s in the 15 Oct run) there are 10 drafts, each with a page, a quote, an explanation, a difficulty and a checker verdict. The host reads, edits or unticks, and adds. Typed-answer drafts also list English-letter spellings, so "Anjana" counts for अंजना.

2. Problem and goals

Goal type	Description
User problem	Writing a show's questions from scanned granths is slow manual work (about 1 hour). The result has no explanations, no deliberate difficulty mix, and typed answers that reject correct spellings.
User goal	Host: a fresh, correct, well-pitched round ready in minutes, with something to teach after every reveal. Player: my right answer counts.
NGO ("business") goal	Run more and fresher teaching shows with the same 1–2 people who host today, and keep every hall engaged from 8 to 75. Spread Jain philosophy accurately; never spread an error.
Product goal	Turn question writing into question <i>reviewing</i> : AI drafts and checks, the host decides. Every added question carries its source, difficulty and explanation. Cost stays at ₹0.

3. AI capability used

- **Multimodal language model** (Google Gemini, flash family, free tier). It reads the **scanned page images** directly. The book has no text layer, and mixes Hindi, Prakrit and Sanskrit.
- **Structured generation**: the model must reply in a fixed JSON schema per round type (multiple choice, typed answer, jumble word or tambola fact), so drafts drop straight into the Studio's own item format.
- **A second, blind evaluation pass** (an "LLM as examinee"). The model answers every drafted question without seeing the answer, and the code compares the two using the site's own answer matcher.

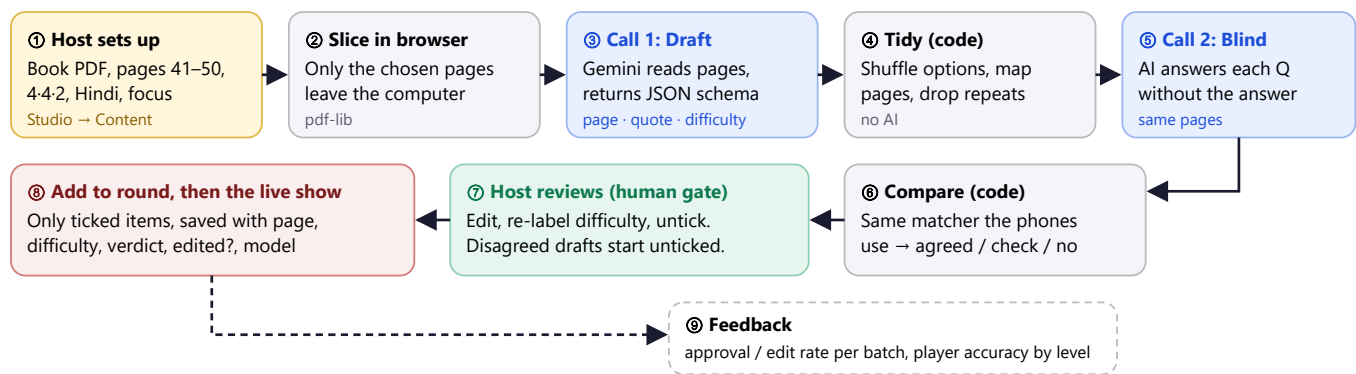
Why this capability fits

The friction is **understanding, not layout**. The work is: read a scanned verse and its commentary, choose a teachable fact, write a fair question and close-but-wrong options, judge difficulty for a family audience, explain it in one sentence, and list how people may spell the answer in English letters. Only a language model can do that. The blind second pass exists because the first prototype's "please verify this answer" check agreed with nonsense (File 03, section 1).

4. Input and output

Input given to AI	AI processing / logic	Output shown to the host
The chosen page range only (≤ 20 pages), cut out of the PDF in the browser	Reads the page images. Skips the cover, contents, index and blank pages. Uses only what is written on these pages.	Drafts grounded in those pages, each with a PDF page number and a quote of ≤ 10 words
Round type and settings: easy/medium/hard counts, language, style ("two clues, one answer"), optional focus topic	Writes exactly that mix in the round's JSON schema. Difficulty rubric is pitched for a family audience; anything that depends on the ऋष distinction is never "easy".	Question, 4 options and the correct one, or an answer plus accepted spellings , or a hint, or a tambola fact. Plus a difficulty label.
Questions already in the round	Avoids repeats. The code also removes duplicates.	Only new questions
The drafts, without their answers (second call)	Answers each question (each clue, for the two-clue style) from the same pages. The code compares answers using the phone's matcher.	Verdict pill: Checker agreed Check this Checker disagreed . Disagreed drafts arrive unticked, with the checker's answer shown.
The host's edits and ticks	None. This is a human decision.	Only ticked drafts are added, in the round's own format, with source metadata

5. AI workflow



Fallback path (calls ③ and ⑤)

Each call walks a list of free "flash" models, newest first and lighter ones last: busy (503), out of quota (429) or retired (404) → next model. A model that does not answer in 120 s counts as busy. If every model is busy, pause 3 s and try the list once more. Then stop with a plain message: "Google's free AI is very busy... You can still add questions by hand or with Add many at once." If only the blind check (⑤) fails, the drafts still arrive, all marked "Check this", because the check is a safety net and should not block the host.

6. Data needs

- **Source book pages** chosen by the host (primary input). Sent to Google Gemini for this request only.
- **Round settings** (type, input mode, difficulty counts, language, style, focus).
- **Existing questions in the round** (text only), so the AI avoids repeating them.
- **Rules and fallbacks** in code: page cap (20), schema, option shuffle, duplicate filter, model list, time-outs.
- **Stored with each added question:** difficulty, PDF page, book name, checker verdict, whether the host edited it, and which model wrote it. This is the evaluation data for File 05.

Sensitive data and how it is handled

- **Player data never leaves Firebase.** Names, phones and ages are not part of any AI request.
- **The API key** is the host's own free key, kept only in that browser (localStorage), never in the code or the database.
- **Books** may carry "for private & personal use". Use stays non-commercial religious teaching, quotes are capped at 10 words, and the pages are not stored by us.
- **Free-tier terms** allow Google to use inputs to improve its products, so the screen says so and private or unpublished manuscripts must not be uploaded.

7. User stories

#	Story
US1	As a host , I want to upload a granth and choose the pages, so that the questions come from exactly the chapter I am teaching this week.
US2	As a host of a mixed-age hall , I want to choose how many easy, medium and hard questions I get, so that children, youth and elders all stay engaged.
US3	As a host , I want every draft to show its page, a line from the book and an independent checker's verdict, so that I can verify it in seconds before it reaches the projector.
US4	As a player without a Hindi keyboard , I want my correct answer typed in English letters to be accepted, so that I am not marked wrong because of my phone.
US5	As a host , I want a short explanation with every question, so that I can teach the idea right after revealing the answer.

#	Story
US6	As a host preparing on show day , I want the tool to try other free models when Google is busy, and to tell me plainly if it cannot, so that I am never stuck without questions.

INVEST check: each story stands alone, is negotiable in its "how", has a clear user value, can be estimated, is small enough for one sprint, and is testable (below).

8. Acceptance criteria (Given-When-Then)

Story	Criterion	Verified
US1	Given a 228-page scanned PDF, when the host picks pages 41–50 and drafts, then only those 10 pages are sent, and every draft cites a page between 41 and 50.	Pass 15 Oct: all 10 drafts cite pp. 44–50
US1	Given a range of more than 20 pages, when the host drafts, then the range is cut to 20 pages from the start page.	Pass code + UI note
US2	Given 4 easy / 4 medium / 2 hard requested, when drafts arrive, then each shows its level, and the host can change the level before adding. In tambola rounds the level becomes the tambola level.	Pass 15 Oct: 4 / 4 / 2 returned
US3	Given a draft whose blind answer differs from the draft's answer, when the review opens, then the draft is unticked, marked "Checker disagreed", and shows the checker's answer.	Pass 14 Oct: the "आयु" draft rejected (checker: "ज्ञान / आयुष्कर्म")
US3	Given any number of drafts, when the host presses Add, then only ticked drafts are added, and nothing is added before that.	Pass 14–15 Oct
US4	Given a typed-answer round, when the AI drafts the answer कषाय, then the accepted answers include English-letter spellings, and a phone that types "Kashaya" is marked correct.	Pass 14 Oct (stored as कषाय Kashaya Kashay)
US5	Given any non-tambola draft, when it arrives, then it has a one- or two-sentence explanation in the chosen language.	Pass after fix (first run returned none)
US6	Given every model answers "busy", when the host drafts, then after two passes they see "Google's free AI is very busy... You can still add questions by hand", and their settings are kept.	Pass 15 Oct, fault-injection test
US6	Given a model that does not reply, when 120 s pass, then the next model is tried automatically.	Pass added after a stuck run on 14 Oct

9. Functional requirements

1. Show " ✨ Draft with AI *Beta*" in Studio → Content, for signed-in hosts only.
2. Accept a PDF and keep it open for the session, so a host can work chapter by chapter without re-uploading.
3. Send only the chosen page range, at most 20 pages.
4. Support Quiz, Jumble, Tambola and Build-your-own rounds that use pick-one or typed answers. Explain politely when a round type is not supported.
5. Let the host set easy, medium and hard counts (0–15 each), language (Hindi or English), style (standard, or two clues and one answer) and an optional focus.
6. Return drafts in the round's own item shape. Typed answers are stored as `main|variant|variant`.
7. Shuffle multiple-choice options so the right answer is not always in the same slot. Pre-select two wrong options for the 50:50 lifeline.
8. Run a blind check on every draft and show a verdict. Untick drafts the checker disagreed with.
9. Let the host edit any field, change the difficulty, and tick or untick before adding. Add only ticked drafts.
10. Save metadata with each added item (difficulty, page, book, verdict, edited, model).

11. Use free models only, with automatic fallback, a 120 s time-out per call, and a clear message on failure.
12. Keep the API key only in the host's browser. Offer a "Get a free key" link and a warning not to use the billed Firebase project.
13. Show the full interface in English and Hindi, matching the rest of the site.

10. Non-functional requirements

Requirement	Target	Observed in testing
Latency	One batch of 10 drafts plus the blind check in ≤ 150 s (P95). The screen explains any wait.	40–149 s across 6 runs (longer when the newest models were busy)
Cost	₹0. Free tier only. No Cloud Functions, no Storage. The Firebase (Blaze) project is never used for AI.	₹0. Paid "pro" models refused the key with a quota error, which proves it has no billing.
Reliability	Works whenever at least one free model is up. 8-model fallback and 2 passes.	Worked through four 503-busy models in one run
Quality gate	100% of added questions are human-approved. 100% are blind-checked when the checker is available.	Enforced in code
Accuracy	$\geq 80\%$ of drafts approved without a change to their facts. To be measured from the saved metadata.	Spot-check of the 15 Oct batch: 8 of 10 usable as written
Privacy & security	Only the chosen pages leave the browser. Host-only screen. The key is stored locally.	Verified: no player data in any request
Accessibility & language	Bilingual interface, large controls, Devanagari output.	Hindi output correct in all runs

11. Assumptions, dependencies and risks

Assumptions

- Hosts have the source books as PDFs (the NGO uses scanned granths)
- Reviewing 10 drafts takes about 10 minutes (to be validated)
- Google keeps a free Gemini tier
- Scanned pages are clear enough to read

Dependencies

- Google Gemini API (free tier) and which models are available
- pdf-lib, loaded from a CDN, for page slicing
- The Studio's existing item formats and answer matcher
- Firebase Hosting (unchanged)

Risks

- A wrong doctrinal "fact" reaches the hall
- Hosts tick everything without reading
- Free quota or busy models on show day
- Book copyright and free-tier data terms
- Differences between traditions

Mitigations are in File 03.

12. Agile backlog

Epic	User story	Priority	Points	Status / notes
AI question drafts	Upload a PDF and slice the page range (US1)	P1	3	Done pdf-lib in the browser
	Draft per round type with a JSON schema (US1)	P1	8	Done 4 round kinds
	Difficulty mix and editable labels (US2)	P1	3	Done
	Blind-answer check and verdict pills (US3)	P1	5	Done replaced the "read-and-agree" check
	English-letter answer variants (US4)	P1	2	Done
	Explanation per question (US5)	P1	1	Done
	Model fallback, time-outs, clear errors (US6)	P1	5	Done
	Review and edit screen; add only ticked drafts (US3)	P1	5	Done
	Re-score past typed answers with the new variants ("shadow scoring")	P2	3	Next feeds the File 05 test
	Monthly scholar spot-check view of AI-written items	P2	3	Next
	Remote on/off switch read from the database	P2	2	Next today rollback is a redeploy (~2 min)

Fibonacci points; 1 point ≈ half a day. 32 points shipped in the first release (14 Oct 2026).